Statistical color models with application to skin det

Jones, M.J. Rehg, J.M.
Res. Lab., Compaq Comput. Corp., Cambridge, MA , USA;

This paper appears in: **Computer Vision and Pattern Recognition, 1999. Computer Society Conference on.**

Meeting Date: 06/23/1999 - 06/25/1999
Publication Date: 23-25 June 1999
Location: Fort Collins, CO USA
On page(s): 280 Vol. 1
Volume: 1
Reference Cited: 15
Number of Pages: 2 vol. (xxiii+637+663)
Inspec Accession Number: 6331104

Abstract:
The existence of large image datasets such as photos on the World Wide Web possible to build powerful generic models for low-level image attributes like c simple **histogram** learning techniques. We describe the construction of **color** skin and non-skin classes from a dataset of nearly 1 billion labeled **pixels**. Th exhibit a surprising degree of separability which we exploit by building a skin detector that achieves an equal error rate of 88%. We compare the performa **histogram** and mixture models in skin detection and find **histogram** models superior in accuracy and computational cost. Using aggregate features compu the skin detector we build a remarkably effective detector for naked people. V this work is the most comprehensive and detailed exploration of skin **color** m date

Index Terms:
image colour analysis image recognition information resources World Wide Web models **histogram** learning image datasets low-level image attributes photos sk skin **pixel** detector statistical **color** models

Home | Log-out | Journals | Conference Proceedings | Standards | Search by Author | Basic Search | Advanced Search | Join IEEE | Web Account | New this week | OPAC Linking Information | Your Feedback | Technical Support | Email Alerting | No Robots Please | Release Notes | IEEE Online Publications | Help | FAQ| Terms | Back to Top

Statistical Color Models with Application to Skin Detection

Michael J. Jones James M. Rehg

Cambridge Research Laboratory
Compaq Computer Corporation
One Kendall Square, Bldg 700
Cambridge, MA 02138
{rehg,mjones}@crl.dec.com

Abstract

The existence of large image datasets such as photos on the World Wide Web make it possible to build powerful generic models for low-level image attributes like color using simple histogram learning techniques. We describe the construction of color models for skin and non-skin classes from a dataset of nearly 1 billion labeled pixels. These classes exhibit a surprising degree of separability which we exploit by building a skin pixel detector that achieves an equal error rate of 88%. We compare the performance of histogram and mixture models in skin detection and find histogram models to be superior in accuracy and computational cost. Using aggregate features computed from the skin detector we build a remarkably effective detector for naked people. We believe this work is the most comprehensive and detailed exploration of skin color models to date.

1 Introduction

A central task in visual learning is the construction of statistical models of image appearance from pixel data. When the amount of available training data is small, sophisticated learning algorithms may be required to interpolate between samples. However, as a result of the world wide web and the proliferation of on-line digital imagery, the vision community today has access to image libraries of unprecedented size and richness. These large data sets can support simple, computationally efficient learning algorithms.

However, a data set such as the web constitutes a biased sample from the space of possible imagery. Thus, the process of building image models from web data must be accompanied by a process of visualizing these models and investigating the statistical characteristics of on-line image data sets. Color is the simplest visual attribute to model, and it is a natural starting point when working with large data sets. Three dimensional color space results in computationally inexpensive algorithms and models that can be visualized easily.

Recently a number of authors have addressed the problem of constructing "generic prior models" [15] of images using multi-scale statistical modeling techniques [4, 15, 1, 11]. In this work, texture models are constructed from the outputs of multi-scale spatial filters, such as wavelets or steerable pyramids. Applications of these models include texture synthesis and classification, as well as noise removal and image coding. In most cases, models are built from a single example image, or a few examples in the case of [1]. A color histogram model can be viewed as the *0th* order version of these spatial models in which the neighborhood structure is limited to a single pixel.

We describe the construction of statistical color models from a data set of unprecedented size: Our model includes nearly 1 billion labeled training pixels obtained from random crawls of the world wide web. From this data we construct a generic color model as well as separate skin and non-skin models. We use visualization techniques to examine the shape of these distributions. We show empirically that the preponderance of skin pixels in web images introduces a systematic bias in the generic distribution of color on the web. We learn both histogram and mixture densities from this data, and show that histogram models slightly outperform mixture models in this domain.

We use skin and non-skin color models to design a skin pixel classifier with an equal error rate of 88%. This is surprisingly good performance given the unconstrained nature of web images. Using our skin classifier, we construct a system for detecting images containing naked people, based on simple aggregate properties of the classifier output. This system compares favorably to recent systems by Forsyth et al. [2] and Wang et al. [13]. This suggests that skin color can be a more powerful cue for detecting people in unconstrained imagery than was previously suspected.

We believe this work is the most comprehensive and detailed exploration of skin color models to date. We are making our labeled dataset of 13,640 photos available to the academic research community. Contact the first author for instructions.

0-7695-0149-4/99 $10.00 © 1999 IEEE

2 Generic Histogram Color Model

There are two issues that must be addressed in building a color histogram model: the choice of color space and the size of the histogram, which is measured by the number of bins per color channel. The 24-bit RGB color space is a natural representation for color images found on the web. High quality color images require 24 bit resolution and images with coarser color quantizations can be mapped into it. In contrast, the size of the histogram depends upon the task. Our starting point is the construction of a histogram model in 24 bit RGB color space. Such a model has a size of 256 bins per color channel, which corresponds to more than 16.7 million (256^3) bins, each mapped to a specific R,G,B color triple. In Section 3.2 we will show that skin classification requires a smaller histogram size for good generalization.

The dataset for the experiments described in this report were obtained by a large crawl of the web which produced about 3 million images (including icons and graphics). A smaller set of images was randomly sampled from this large set and cleared of all icons and graphics by hand. This produced a set of 18,696 photographs. This is a dataset of nearly 2 billion pixels, which is two orders of magnitude more data than the number of degrees of freedom in a histogram model of size 256. We used this data to construct a generic histogram color model.

The counts in the histogram bins are converted to a discrete probability distribution $P(\cdot)$ in the usual manner:

$$P(rgb) = \frac{c[rgb]}{T_c}, \qquad (1)$$

where $c[rgb]$ gives the count in the histogram bin associated with the RGB color triple rgb and T_c is the total count obtained by summing over all of the bins in the histogram.

To visualize this probability distribution, we display the histogram as a 3-D model in which each bin is rendered as a cube whose size is proportional to the number of counts it contains. The color of each cube corresponds to the smallest enclosed RGB triple. Figure 1 shows a sample view of the histogram which was produced by our visualization tool. This rendering uses a perspective projection model with a viewing direction along the green-magenta axis which joins corners $(0, 255, 0)$ and $(255, 0, 255)$ in color space. The histogram in Figure 1 has 8^3 bins and only shows bins with counts greater than 336,818. Downsampling and thresholding the full size model makes the global structure of the distribution more apparent.

By examining the 3-D histogram from several angles its overall shape can be inferred. Another visualization of the model can be obtained by integrating the 3-D density along the viewing direction and plotting the resulting 2-D marginal density function as a surface. Figure 2 shows

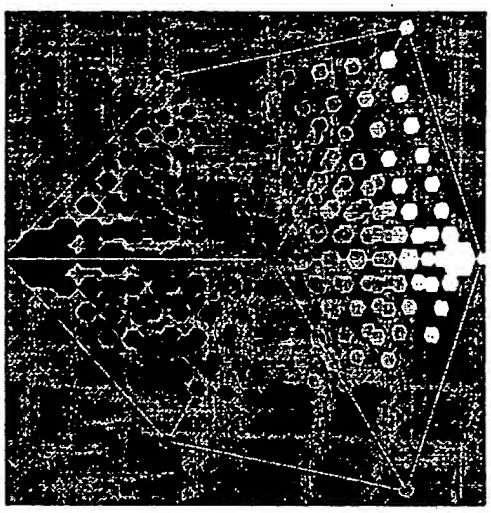

Figure 1: 3-D full color histogram model viewed along the green-magenta axis.

the marginal distribution that results from integrating along the same green-magenta axis used in Figure 1. The positions of the black-red and black-blue axes under projection are also shown. The density is concentrated along a ridge which follows the gray line[1] from black to white. White has the highest likelihood, followed closely by black.

Additional information about the shape of the surface in Figure 2 can be obtained by plotting its equiprobability contours. These are shown in Figure 3. This plot reinforces the conclusion that the density is concentrated around the gray line and is more sharply peaked at white than black. An intriguing feature of this plot is the bias in the distribution towards red. In the next section, we will demonstrate empirically that this bias is due largely to the presence of skin in web images.

It is also interesting to note that in gathering our data set we have found that 77% of the possible colors are never encountered (i.e. the histogram is mostly empty), and about 52% of web images have people in them.

3 Skin and Non-skin Color Models

Our next step is to specialize the generic color model into separate skin and non-skin models using labeled training pixels. These models can be used for skin detection in color images. The color of skin in images depends primarily on the amount of hemoglobin and melanin in the dermis and on the conditions of illumination. It is well-known that the hue of skin is roughly invariant across different ethnic

[1]The gray line is the projection of the gray axis which connects the black $(0, 0, 0)$ and white $(255, 255, 255)$ corners of the cube.

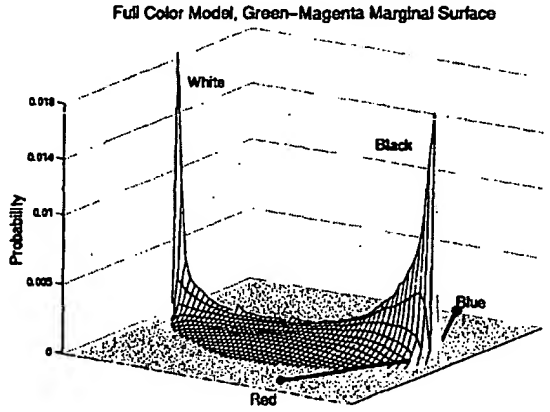

Figure 2: Surface plot of the marginal density formed by integrating along the viewing direction in Figure 1.

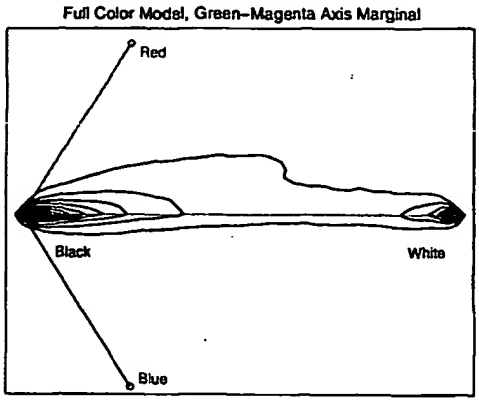

Figure 3: Equiprobability contours from the surface plot in Figure 2.

groups after the illuminant has been discounted. This is because differences in skin color result primarily from differences in the concentration of melanin, which affects the intensity of skin color but not its hue.

Unfortunately we do not know the illumination conditions in an arbitrary image[2] and so the variation in skin colors is much less constrained in practice. This is particularly true for web images captured under a wide variety of conditions. However, given a large collection of labeled training pixels we can still model the distribution of skin and non-skin colors in un-normalized color space.

We constructed skin and non-skin histogram models using a subset of 13,640 photos sampled from the total set of 18,696 photographs described in Section 2. In the 4675 photos containing skin, the skin pixels were segmented by hand using a tool which allowed a person to carefully "paint" the skin region. Areas of the face such as eyes, teeth and hair were not labeled as skin. Pixels not labeled as skin in these photos were discarded to reduce the chance that segmentation errors would contaminate the model. These labeled skin pixels were placed into the skin histogram model. The remaining 8965 photos which did not contain any skin were placed into the non-skin model. These two models together contain almost one billion labeled pixels, which includes more than 38.7 million hand labeled skin pixels! In passing we note that skin pixels make up about 10% of the total pixels in our dataset.

Given these two histograms, we can compute the probability that a given pixel belongs to the skin and non-skin classes:

$$P(rgb|skin) = \frac{s[rgb]}{T_s}, \qquad P(rgb|\neg skin) = \frac{n[rgb]}{T_n} \quad (2)$$

where $s[rgb]$ is the pixel count contained in bin rgb of the skin histogram, $n[rgb]$ is the equivalent count from the non-skin histogram, and T_s and T_n are the total counts contained in the skin and non-skin histograms, respectively.

The skin and non-skin color models can be examined using the same techniques we employed with the generic color model. Contour plots for marginalizations of the skin and non-skin models are shown in Figures 4 and 5. They are formed by integrating along the same green-magenta axis used in Figure 3. These plots show that a significant degree of separation exists between the skin and non-skin models. The non-skin model is concentrated along the gray axis, while the majority of the probability mass in the skin model lies off this axis. This separation between the two classes is the basis for the good performance of our skin classifier, described in Section 3.1.

It is interesting to compare the non-skin color model illustrated in Figure 5 with the full color model shown in Figure 3. The only difference in the construction of these two models is the absence of skin pixels in the non-skin case. Note that the result of omitting skin pixels is a remarkable increase in the symmetry of the distribution around the gray axis. This observation suggests that although skin pixels constitute only about 10% of the total pixels in the dataset, they exhort a disproportionately large effect on the shape of the generic color distribution for web images, biasing it strongly in the red direction. We suspect that this effect results from the fact that the skin class occurs more frequently than other classes of object colors

[2]The illuminant could be discounted, however, if a solution to the color constancy problem were available.

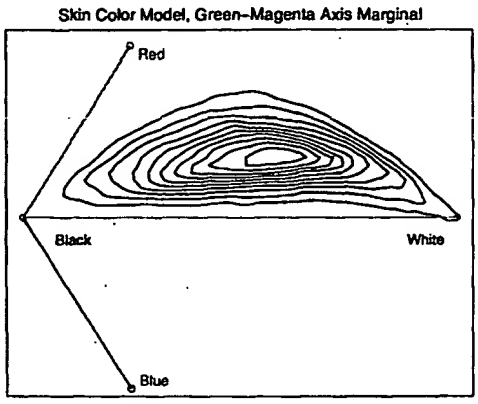

Figure 4: Contour plot of the skin color model integrated along the same viewing direction as in Figure 3.

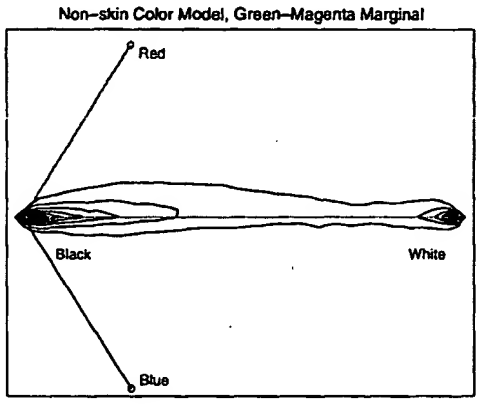

Figure 5: Contour plot of the non-skin color model integrated along the same viewing direction as in Figure 3.

(52% of our images contained skin). See [6] for more details, including additional visualizations.

3.1 Skin Pixel Detection

Given the skin and non-skin distributions, we can obtain a pixel classifier through the standard likelihood ratio approach [3]. A particular RGB value is labeled skin if

$$\frac{p(rgb|skin)}{p(rgb|\neg skin)} \geq \Theta, \qquad (3)$$

where $0 \leq \Theta \leq 1$ is a threshold. Θ depends upon the application-specific costs of classification errors, as well as on the prior probabilities of skin and non-skin, $P(skin)$ and $P(\neg skin)$. One reasonable choice of prior is $P(skin) = T_s/(T_s + T_n)$.

An important property of equation 3 is the receiver operating characteristic (ROC) curve, which shows the relationship between correct detections and false detections as a function of the detection threshold Θ [12]. The ROC curve provides a global measure of classifier performance which can be used to compare classifier designs. It is also a useful tool when setting detection thresholds for a particular application.

In order to test our classifier, we divided our labeled pixel data into separate training and testing sets. The test set consisted of 2336 skin images and 4482 non-skin images taken from the set of 13,640 labeled photos. The ROC curve for the skin classifier on this test data is shown as plot 5 (the topmost curve) in figure 6. The axis "probability of correct detections" gives the fraction of pixels labeled as skin that were classified correctly, while "probability of false detections" gives the fraction of non-skin pixels which are mistakenly classified as skin.

The performance of the classifier, as measured by the ROC curve, is surprisingly good given the unconstrained nature of web imagery. The classifier has an equal error rate of 88%. This corresponds to the point on the ROC curve where the probability of false rejection (which is one minus the probability of correct detection) equals the probability of false detection. The area under the ROC curve is 0.942 (it would be 1.0 for a perfect detector). Both the equal error rate and the area under the ROC curve provide scalar measures of overall classifier performance.

Figure 7 shows some representative examples of the skin classifier's performance (with $\Theta = 0.4$) on images from our test set. Directly below each image is a mask image in which detected skin pixels are drawn in black. The classifier does a good job of detecting skin in most of these examples, but tends to fail on either highly saturated or shadowed skin. In many of the non-skin images the false detections are sparse and scattered (e.g. the flowers image). More problematic are images with wood or copper-colored metal which are hard to discriminate from skin (e.g. the railroad tracks image).

Note that the use of color spaces other than RGB (such as YUV or HSV) will not improve the performance of the skin detector. Detector performance depends entirely on the amount of overlap between the skin and non-skin samples. Colors which occur in both the skin and non-skin classes with comparable frequencies cannot be classified reliably. No fixed global transformation between color spaces can affect the degree of overlap.

3.2 Analysis of Skin Model

It is interesting to examine the effect of modeling decisions on the classifier performance. Here we focus on three factors: The amount of training data, the number of

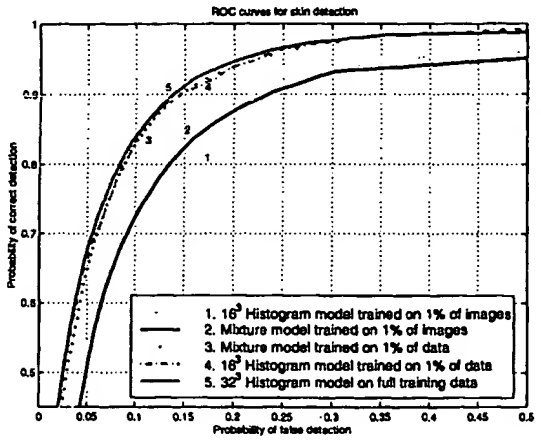

Figure 6: ROC curves for a family of skin detectors based on different histogram and mixture modeling choices. The best ROC curve (number 5) is the result of using a 32^3 bin histogram model.

Figure 7: Examples of skin pixel classifier performance.

histogram bins, and a comparison to mixture of Gaussian models. We can summarize our findings as follows: 1) The large size of our data set is crucial, 2) Histograms with 32 bins/channel give the best performance, suggesting that significant generalization is required, 3) Histogram models are superior to mixture of Gaussian models, 4) Using a small amount of data sampled from a large dataset also produces good results. Each of these points is discussed below, see [6] for the complete details.

We constructed a series of histogram color models with varying amounts of training data. This resulted in a family of ROC curves indexed by the number of training pixels. Our dataset of 13,640 labeled photos represents the empirical limiting point of this progression, at which adding additional pixels did not significantly improve performance. The importance of a large dataset is underscored by ROC curve no. 1 in Figure 6. This classifier was constructed from 1% of the available training images and exhibits relatively poor performance (the ROC curve area is 0.890, compared to 0.942 for the full data model).

Generalization in the histogram model is controlled by the number of bins. We found that a histogram with 32^3 bins (for both skin and non-skin) performed the best when using our full dataset. The ROC curve for this classifier is shown as plot 5 (the topmost curve) in Figure 6. Increasing or decreasing the number of bins reduced the performance (see [6] for details.)

An alternative form of generalization is provided by mixture of Gaussian models. Mixture models have been popular in earlier skin color modeling work [5, 14] and we examined their performance on our dataset. Using the EM algorithm [8], we fit separate mixture models with 16 Gaussian each to the full set of skin and non-skin pixel data. The performance of the resulting classifier was slightly worse than the 32 bin histogram model. Its ROC curve area was 0.932 in comparison to 0.942.

Mixture models might be expected to do better as the size of the state space increases or as the amount of training data decreases. We tested this hypothesis by fitting mixtures to the reduced dataset containing 1% of the total images. The ROC curve for the resulting mixture model is number 2 in Figure 6. Its area is 0.895, compared to 0.890 for the histogram model on this dataset, a slight improvement. However, since mixture models are computationally more expensive than histograms during both learning and evaluation, these results suggest that histograms are the

best choice for skin color modeling.

Finally, we tested the performance of the models trained on a small set of data sampled uniformly from the large training set. We sampled 387,172 skin pixels and 4,261,703 non-skin pixels (1% of the training data) and built both histogram and mixture models from this data. The performance of these models is also shown in figure 6. The area under the ROC curve for the histogram model trained on 1% of the data is 0.9405 and for the mixture model it is 0.9378. They are almost as good as the histogram model using the full training set. This demonstrates that while a large data set is necessary to capture the underlying distribution of skin and non-skin colors, it is sufficient to train models on a smaller set of samples.

4 Adult Image Detection

By taking advantage of the fact that there is a strong correlation between images with large patches of skin and adult or pornographic images, the skin detector can be used as the basis for an adult image detector. The ability to filter out adult images is important for image search engines on the web that wish to avoid offensive content.

To detect adult images, a feature vector is formed based on the output of the skin detector and then a neural network classifier is trained on a set of labeled feature vectors. The features we used are:

- Percentage of pixels detected as skin
- Average probability of the skin pixels
- Size of the largest connected component of skin
- Number of connected components of skin
- Percent of novel pixels (those with zero counts in both the skin and non-skin histograms)
- height of the image
- width of the image

We used 10681 images which were manually classified into adult and non-adult sets to train a neural network classifier. The neural network outputs a number between 0 and 1 with 1 signifying an adult image. We can threshold this value to make a binary decision. By varying the threshold, we get the ROC curve shown in figure 8 for the training data.

To test the adult image detector, we gathered images from two new crawls of the web. Crawl A used adult sites as starting points for the crawl and so gathered many adult images. Crawl B used non-adult sites as starting points and gathered very few adult images. Crawl A consisted of 2365 HTML pages containing 5241 adult images and 6082 non-adult images (including icons and other graphics). Crawl B consisted of 2692 HTML pages containing 3 adult images and 13970 non-adult images. The classification for each image was again determined manually.

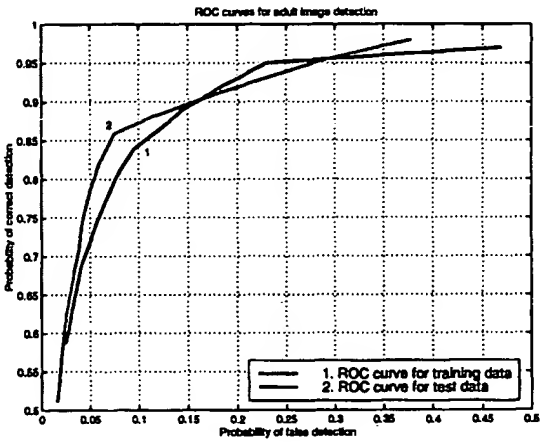

Figure 8: ROC curves for the adult image detector on both training and testing images.

The important statistics for the popular detector on these test sets is the percentage of correct detections for the set of adult images from crawl A and the percentage of false positives for the set of non-adult images from crawl B. The ROC curve for the adult image detector for this training data is also shown in figure 8. The performance is very good considering color is the only feature used. For example, the classifier attains about 85.8% correct detections with about 7.4% false positives.

We have also explored combining the adult image detector just described with a text-based classifier which uses the text around an image on an HTML page to determine if an image is pornographic. The combined detector correctly labels 93.9% of the adult images from crawl A and obtains 8% false positives on the non-adult images from crawl B. The text-based detector by itself correctly labels 84.9% of the adult images with 1.1% false positives.

The results show that simply analyzing color values allows very good detection of adult images. Not surprisingly, adding information from the surrounding text can boost performance significantly.

5 Previous Work

There have been a number of researchers who have looked at using color information to detect skin. We believe we are the first to build a general statistical model of skin color from a large data set. Forsyth et. al. [2] and Rowley et. al. [9] have employed ad hoc skin color models as a preprocessor in analyzing large image databases. Other researchers have built small scale skin models using single Gaussians [14], mixtures of Gaussians [5] or histograms [10, 7]. Most of these models are based on skin data ac-

quired from a limited number of people under a limited range of lighting conditions. Our work demonstrates the superiority of histogram models: They are equivalent to mixture models in accuracy and are more efficient computationally.

Forsyth et. al. [2] and Wang et. al. [13] have also looked at the problem of detecting adult images. Both used a simple color model and emphasized shape and texture cues. In contrast we have used a more accurate color model to construct simple spatial features. It is interesting that our detection results are comparable to theirs. This suggests that color is a more powerful cue than might have been expected.

6 Conclusions

The existence of large image datasets such as the photos on the World Wide Web make it possible to build powerful generic models for low-level image attributes like color using simple histogram learning techniques. We have demonstrated this point empirically by constructing generic color models, as well as specialized skin and non-skin color models, from nearly 1 billion labeled pixels. We are making our labeled dataset of 13,640 photos available to the academic research community. Contact the first author for instructions.

We demonstrate that a significant degree of separability exists between the skin and non-skin distributions, which we exploit in building a skin pixel detector with an equal error rate of 88%. Furthermore, we show empirically that the preponderance of skin pixels in Web images leads to a systematic bias in the generic distribution of color on the Web. We explore the performance of both histogram and mixture of Gaussian models in classification, and find histogram models to be superior in accuracy and speed. We believe this work to be the most comprehensive and detailed exploration of skin color models to date.

We demonstrate a surprisingly effective detector for images containing naked people which is based on the output of our skin pixel classifier. This suggests that skin color can be a more powerful cue for detecting people in unconstrained imagery than was previously suspected.

Acknowledgments

The authors would like to thank Michael Swain and Henry Schneiderman for some valuable discussions. We would also like to thank Pedro Moreno for his help in fitting the mixture models using a parallel implementation of the EM algorithm. Thanks to Nick Whyte of AltaVista for providing the image dataset.

References

[1] J. S. De Bonet and P. Viola. Texture recognition using a non-parametric multi-scale statistical model. In *Proc. Computer Vision and Pattern Recognition*, pages 641–647, 1998.

[2] D. A. Forsyth, M. Fleck, and C. Bregler. Finding naked people. In *Proc. Fourth European Conference on Computer Vision*, pages 593–602, 1996.

[3] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 1972.

[4] D. J. Heeger and J. R. Bergen. Pyramid-based texture analysis/synthesis. In *SIGGRAPH '95*, pages 229–238, 1995.

[5] T. S. Jebara and A. Pentland. Parameterized structure from motion for 3d adaptive feedback tracking of faces. In *Proc. Computer Vision and Pattern Recognition*, pages 144–150, 1997.

[6] M. J. Jones and J. M. Rehg. Statistical color models with application to skin detection. Technical Report CRL 98/11, Compaq Cambridge Research Lab., 1998.

[7] R. Kjeldsen and J. Kender. Finding skin in color images. In *Face and Gesture (FG96)*, pages 312–317, 1996.

[8] R. Redner and H. Walker. Mixture densities, maximum likelihood, and the EM algorithm. *SIAM Revew*, 26:195–239, 1994.

[9] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. In *Proc. Computer Vision and Pattern Recognition*, pages 203–208, 1996.

[10] B. Schiele and A. Waibel. Gaze tracking based on face-color. In *Face and Gesture (FG95)*, pages 344–349, 1995.

[11] E. P. Simoncelli. Statistical models for images: Compression, restoration and synthesis. In *31st Asilomar Conf. on Sig., Sys. and Comp.*, 1997.

[12] H. Van Trees. *Detection, Estimation, and Modulation Theory*, volume I. Wiley, 1968.

[13] J. Z. Wang, J. Li, G. Wiederhold, and O. Firschein. System for screening objectionable images using daubechies' wavelets and color histograms. In *Proc. IDMS*, 1997.

[14] J. Yang, W. Lu, and A. Waibel. Skin-color modeling and adaptation. In *Proc. ACCV*, pages 687–694, 1998.

[15] S. C. Zhu, Y. Wu, and D. Mumford. Filters, random fields and maximum entropy. *Intl. J. of Computer Vision*, 27(2):107–126, March 1998.

US-PAT-NO: 6272239

DOCUMENT-IDENTIFIER: US 6272239 B1
See image for Certificate of Correction

TITLE: Digital image color correction device and method
employing fuzzy logic

DATE-ISSUED: August 7, 2001

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Colla; Federica	Crema	N/A	N/A	IT
Mancuso; Massimo	Monza	N/A	N/A	IT
Poluzzi; Rinaldo	Milan	N/A	N/A	IT

APPL-NO: 09/ 222247

DATE FILED: December 28, 1998

FOREIGN-APPLICATION-PRIORITY-DATA:

A digital color correction device and

digital video image
and computing a multilevel value representing a membership of each pixel to a
skin color class; a global parameter estimator (2) receiving in input each of
said pixel and the relative membership value, and computing a first and a
second parameter which define the characteristics of a portion of said image
that belongs to said skin color class; a processing unit (3) connected
downstream to said global parameter estimator and to said pixel fuzzifier unit
and adapted to correct each of the pixels of said portion of the image that
belongs to said skin color class, according to said first global parameter
(300), to obtain corrected pixels; and a processing switch (4) for outputting
said pixels or said corrected pixels according to said second global parameter
(400).

29 Claims, 9 Drawing figures

Exemplary Claim Number: 1

require.

Claims Text - CLTX (34):

15. The method according to claim 14, wherein the area of the skin **color class** is evaluated on the basis of the **percentage of pixels** belonging to the **class**.

Fuzzy Color Histogram and Its Use in Color Image Retrieval

Ju Han and Kai-Kuang Ma, *Senior Member, IEEE*

***Abstract*—A conventional color histogram (CCH) considers neither the color similarity across different bins nor the color dissimilarity in the same bin. Therefore, it is sensitive to noisy interference such as illumination changes and quantization errors. Furthermore, CCHs large dimension or histogram bins requires large computation on histogram comparison. To address these concerns, this paper presents a new color histogram representation, called *fuzzy color histogram* (FCH), by considering the color similarity of each pixel's color associated to all the histogram bins through fuzzy-set membership function. A novel and fast approach for computing the membership values based on fuzzy c-means algorithm is introduced. The proposed FCH is further exploited in the application of image indexing and retrieval. Experimental results clearly show that FCH yields better retrieval results than CCH. Such computing methodology is fairly desirable for image retrieval over large image databases.**

***Index Terms*—Conventional color histogram, fuzzy c-means, fuzzy color histogram, illumination changes, image indexing and retrieval, membership matrix.**

I. Introduction

NUMEROUS methods about efficient image indexing and retrieval from image databases have been proposed for the applications such as digital library [1]–[3]. Low-level visual features such as color, texture, and shape are often employed to search relevant images based on the query image. Among these features, color constitutes a powerful visual cue and is one of the most salient and commonly used features in color image retrieval systems.

Swain and Ballard [4] have demonstrated the potential of using color histograms for color image indexing. Because each histogram bin represents a local color range in the given color space, color histogram represents the coarse distribution of the colors in an image. Two similar colors will be treated as identical provided that they are allocated into the same histogram bin. On the other hand, two colors will be considered totally different if they fall into two different bins even though they might be very similar to each other. This makes color histograms sensitive to noisy interference such as illumination changes and quantization errors. In this paper, we proposed a new color histogram, called *fuzzy color histogram* (FCH), to efficiently address the aforementioned issue.

In contrast with conventional color histogram (CCH) which assigns each pixel into one of the bins only, our FCH considers the color similarity information by spreading each pixel's total membership value to all the histogram bins. Furthermore, to save computation, we introduce an efficient method to compute these membership values using *fuzzy c-means* (FCM) clustering algorithm. Experimental results show that the obtained FCH is less sensitive to noisy interference such as lighting intensity changes and quantization errors than CCH.

Moreover, in contrast with quadratic histogram distance exploited for measuring the degree of similarity between CCHs, simple Euclidean distance measurement over their FCHs can yield similar retrieval results. This is a fairly attractive and desirable computing paradigm for the application of image indexing and retrieval especially over large image databases.

In the next section, we introduce related works of color histogram based methods for image indexing and retrieval. The concept of FCH is introduced in Section III. An efficient scheme to compute the required fuzzy membership values using FCM algorithm is introduced in Section IV. In Section V, we analyze the relationship between FCH and other color histograms. In Section VI, we analyze the experimental results of image retrieval based on FCH and discuss the parameter selection of FCH. Section VII concludes the paper.

II. Related Works

Color histograms are easy to compute, and they are invariant to the rotation and translation of image content. However, color histograms have several inherent problems for the task of image indexing and retrieval. The first concern is their sensitivity to noisy interference such as lighting intensity changes and quantization errors. The second problem is their high dimensionality on representation. Even with coarse quantization over a chosen color space, color histogram feature spaces often occupy more than one hundred dimensions (i.e., histogram bins) [5] which significantly increases the computation of distance measurement on the retrieval stage. Finally, color histograms do not include any spatial information and are therefore incompetent to support image indexing and retrieval based on local image contents. In the following, we briefly describe several existing approaches that have been attempting to address these concerns.

A. Sensitivity

Some approaches exploit the color histogram derived together with a similarity measurement chosen to make color

Manuscript received August 23, 2000; revised March 4, 2002. This work was published in part in the *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, June 2000. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jean-Luc Dugelay.

J. Han is with the Department of Electrical Engineering, University of California, Riverside, CA 92521 USA (e-mail: jhan@vislab.ucr.edu).

K.-K. Ma is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: ekkma@ntu.edu.sg).

Publisher Item Identifier 10.1109/TIP.2002.801585.

histograms more robust to noisy interference. To identify objects based on their color histograms, Swain and Ballard [4] propose a *histogram intersection* method which is able to eliminate the influence of color contributed from the background pixels during the matching process in most cases. Although their method is robust to object occlusion and image resolution, but it is still sensitive to illumination changes [4].

Funt and Finlayson [6] propose a *color constant color indexing* method to extend Swain and Ballard's color indexing method to be illumination independent by establishing the histogram of color ratios. Since the illumination remains essentially constant locally, calculating the ratios of neighboring colors removes the illumination variation component. Similar extension can be found in Drew *et al.*'s work [7].

Cumulative color histogram [8] utilizes the spatial relationship of the histogram bins in the color space. Consequently, it is slightly more robust with respect to illumination changes than CCH [8]. In Section V, we will show that it can be viewed as a special case of our FCH.

QBIC [1] takes into account the perceptual color similarity between histogram bins through the measurement of *quadratic distance*, which is a weighted distance between two CCHs with each weight denoting the similarity between a pair of color histogram bins. It has been shown that such measurement is more closely related to human being's judgment on color similarity comparison, but on the expense of large computations.

B. Dimensionality

Many other approaches exploit their derived color histogram methods to facilitate the design of efficient database indexing schemes. Hafner *et al.* [9] generalize computationally simple similarity measures using *singular value decomposition* (SVD) method to compute quadratic histogram distance. It has been mathematically shown that SVD-based approach provides the lower bounds on the histogram distance measure. Mandal *et al.* [10] reduce the computational complexity of color histogram comparison by representing the histogram in terms of its moments. Experimental results also indicate that Legendre moments provide superior retrieval performance compared to regular moments [10].

C. Spatial Information

Some approaches strive to incorporate spatial information into color histograms by dividing each image into subregions and imposing positional constraints on image comparison in order to increase image discrimination power [11]–[14]. Smith and Chang's method [11] uses back projection of binary color sets to extract color regions. Each of these regions is efficiently represented by a binary color set and its location information as well. Stricker and Dimai's method [12] tessellates each image into five partially overlapping fuzzy regions and extracts the first two color moments of each region both weighted by the membership functions of the region, respectively, to form a feature vector for the image.

Other approaches augment histograms with local spatial properties. Pass and Zabih [15] propose a split histogram, called *color coherence vector* (CCV), where image pixels in a given histogram bin are partitioned into two classes based on their spatial coherence [16]. A pixel is considered as coherent pixel if it is part of a sizable contiguous region; otherwise, incoherent pixel. Huang *et al.* [17], [18] propose *color correlograms* to take into account the local color spatial correlation as well as the global distribution of this spatial correlation. In fact, a color correlogram of an image forms a table of statistics for color pairs, where the k-th entry for pair $\langle i, j \rangle$ specifies the probability of finding a pixel of color j from a pixel of color i at a distance k in the image.

All the above-mentioned approaches made some improvements over the CCH for the task of image indexing and retrieval. Our FCH proposed in this paper makes improvement on robustness (less sensitive to interference), efficiency (reduced dimension), and computation (less online computation consumed). The full development of FCH is presented as follows.

III. Fuzzy Color Histogram

In this paper, the color histogram is viewed as a color distribution from the probability viewpoint. Given a color space containing n color bins, the color histogram of image I containing N pixels is represented as $H(I) = [h_1, h_2, \ldots, h_n]$, where $h_i = N_i/N$ is the probability of a pixel in the image belonging to the ith color bin, and N_i is the total number of pixels in the ith color bin. According to the total probability theory, h_i can be defined as follows:

$$h_i = \sum_{j=1}^{N} P_{i|j} P_j = \frac{1}{N} \sum_{j=1}^{N} P_{i|j} \tag{1}$$

where P_j is the probability of a pixel selected from image I being the jth pixel, which is $1/N$, and $P_{i|j}$ is the conditional probability of the selected jth pixel belonging to the ith color bin.

In the context of CCH, $P_{i|j}$ is defined as

$$P_{i|j} = \begin{cases} 1, & \text{if the } j\text{th pixel is quantized into the } i\text{th color bin} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

This definition leads to the boundary issue of CCH such that the histogram may undergo abrupt changes even though color variations are actually small. This reveals the reason why the CCH is sensitive to noisy interference such as illumination changes and quantization errors.

The proposed FCH essentially modifies probability $P_{i|j}$ as follows. Instead of using the probability $P_{i|j}$, we consider each of the N pixels in image I being related to all the n color bins via fuzzy-set membership function such that the degree of "belongingness" or "association" of the jth pixel to the ith color bin is determined by distributing the membership value of the jth pixel, μ_{ij}, to the ith color bin.

Definition (Fuzzy Color Histogram): The fuzzy color histogram (FCH) of image I can be expressed as $F(I) = [f_1, f_2, \ldots, f_n]$, where

$$f_i = \sum_{j=1}^{N} \mu_{ij} P_j = \frac{1}{N} \sum_{j=1}^{N} \mu_{ij}. \tag{3}$$

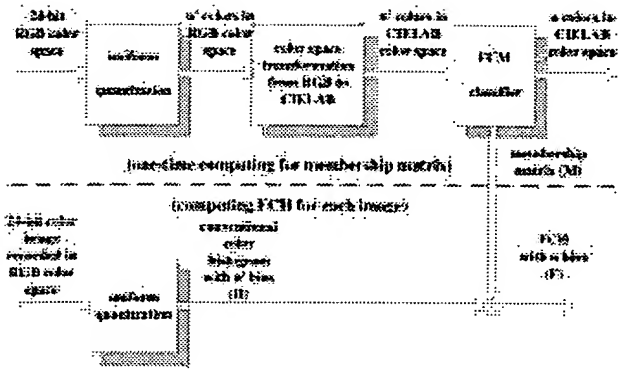

Fig. 1. Procedure diagram for computing FCH ($n' = 16^3 = 4096$ in our experiment).

P_j has been defined in (1), and μ_{ij} is the membership value of the jth pixel in the ith color bin.

In contrast with CCH, our FCH considers not only the similarity of different colors from different bins but also the dissimilarity of those colors assigned to the same bin. Therefore, FCH effectively alleviates the sensitivity to the noisy interference.

IV. FCH Computing

Equation (3) gives the definition of FCH, but it does not provide an applicable method to compute FCH. Given two colors i and j, Hafner *et al.* [9] measure their perceptual similarity in terms of the Euclidean distance between colors i and j represented in a chosen color space. However, the measurement does not consider the nonuniformity inherent in color space representation. To accurately quantify the perceptual color similarity between two colors recorded in a specific color space, the nonuniformity of that color space should be considered. For that, we choose the CIELAB color space which is one of perceptually uniform color spaces and has been increasingly exploited into many electronic color imaging systems (e.g., Postscript language and Adobe Photoshop) [19].

Since RGB color space has been most commonly used for representing color images, intuitively we need to perform nonlinear color space transformation from RGB to CIELAB pixel by pixel. Such pixel-wise transformation is computationally intensive for the entire image. Moreover, to compute the FCH of a color image, we need to compute each pixel's membership values with respect to all available color bins, respectively. Such direct approach is also not favorable because of its large computational load. To address the above-mentioned issues, we propose an efficient method to compute FCH based on *fuzzy c-means* (FCM) clustering algorithm [20]. The procedure diagram for computing FCH is illustrated in Fig. 1.

First, we perform fine uniform quantization in RGB color space by mapping all pixel colors to n' histogram bins. Here, the bin number n' is chosen to be large enough so that it makes the color difference between two adjacent bins small enough. Then, we transform the n' colors from RGB to CIELAB color space. Finally, we classify these n' colors in CIELAB color space to n clusters using FCM clustering technique (usually, $n \ll n'$; hence, a coarse quantization process), with each cluster representing an FCH bin. Through these steps, a pixel's membership value to an FCH bin can be represented by the corresponding fine color bin's membership value to the coarse color bin. Note that we only need to compute these membership values once, and they are represented as a membership matrix $M = [m_{ij}]_{n \times n'}$. Each element m_{ij} in M is the membership value of the jth fine color bin distributing to the ith coarse color bin. Thus, the FCH of an image can be directly computed from its CCH without computing membership values for each pixel. That is, given an n'-bin CCH $H_{n' \times 1}$, the corresponding n-bin FCH $F_{n \times 1}$ can be computed as follows:

$$F_{n \times 1} = M_{n \times n'} H_{n' \times 1} \tag{4}$$

where membership matrix M is pre-computed only once and can be used to generate FCH for each database image. We employ FCM clustering algorithm to not only classify the n' fine colors to n clusters but also obtain membership matrix M at the same time. For the latter, we explain how it works with more details as follows.

FCM is an unsupervised clustering algorithm that has been applied successfully to a number of problems involving feature analysis, clustering and classifier design. The FCM minimizes an objective function J_m, which is the weighted sum of squared errors within each group, and is defined as follows [21]:

$$J_m(U, V; X) = \sum_{k=1}^{n} \sum_{j=1}^{c} u_{ik}^m \|x_k - v_i\|_A^2, \quad 1 < m < \infty \tag{5}$$

where $V = [v_1, v_2, \ldots, v_c]^T$ is a vector of unknown cluster prototypes. The value of u_{ik} represents the membership of the data point x_k from the set $X = \{x_1, x_2, \ldots, x_n\}$ with respect to the ith cluster. The inner product defined by a norm matrix A defines a measurement of similarity between a data point and the cluster prototypes, respectively. A nondegenerate fuzzy c-partition of X is conveniently represented by a matrix $U = [u_{ik}]$. The weighting exponent m controls the extent of membership shared by c clusters.

It has been shown by Bezdek [20] that if $\|x_k - v_i\|_A > 0$ for all i and k and $m > 1$, then J_m could be minimized at (U, V) where

$$v_i = \frac{\sum_{k=1}^{n} (u_{ik})^m x_k}{\sum_{k=1}^{n} (u_{ik})^m}, \quad \text{for } 1 \le i \le c, \tag{6}$$

and

$$u_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{\|x_k - v_i\|_A^2}{\|x_k - v_j\|_A^2} \right)^{\frac{1}{m-1}}}, \quad \text{for } 1 \le i \le c \text{ and } 1 \le k \le n. \tag{7}$$

Equations (6) and (7) cannot be solved analytically, but an approximate solution can be obtained by performing the following iterative procedures. (First, denote (l) as the iteration index.)

Algorithm (Fuzzy C-Means):

Step 1) Input the number of clusters c, the weighting exponent m, and error tolerance ϵ.
Step 2) Initialize the cluster centers v_i, for $1 \le i \le c$.
Step 3) Input data $X = \{x_1, x_2, \ldots, x_n\}$.
Step 4) Calculate the c cluster centers $\{v_i^{(l)}\}$ by (6).
Step 5) Update $U^{(l)}$ by (7).

Step 6) If $\|U^{(l)}-U^{(l-1)}\| > \epsilon$, $l = l+1$ and return to Step 4; otherwise, stop.

In our work, we need to classify the n' fine colors in CCH into n clusters for FCH. Due to the perceptual uniformity of CIELAB color space, the inner product $\|x_k - v_i\|_A^2$ can be simply replaced by $\|x_k - v_i\|^2$, which is the Euclidean distance between the fine color x_k and the cluster center v_i. The fuzzy clustering result of FCM algorithm is represented by matrix $U = [u_{ik}]_{n\times n'}$, and u_{ik} is referred to as the grade of membership of color x_k with respect to cluster center v_i. Thus, the obtained matrix $U_{n\times n'}$ can be viewed as the desired membership matrix $M_{n\times n'}$ for computing FCH, i.e., $M_{n\times n'} = U_{n\times n'}$. Moreover, the weighting exponent m in FCM algorithm controls the extent or "spread" of membership shared among the fuzzy clusters. Therefore, we can use the parameter m to control the extent of similarity sharing among different color bins in FCH. The membership matrix M can be thus adjusted according to different image retrieval applications. In general, if higher noisy interference is involved, larger m value should be used.

V. Relationship Between FCH and Other Color Histograms

Cumulative color histogram [8] has been proven to be more robust to noisy interference than CCH. Given the color histogram H of image I, the corresponding cumulative color histogram is mathematically represented as $\check{H}(I) = [\check{h}_1, \check{h}_2, \ldots, \check{h}_n]$, where $\check{h}_i = \sum_{C_j \le C_i} h_j$. Here, C_i and C_j are the representative color values of the ith and jth histogram bins, respectively. In RGB color space, $C_j = (r_j, g_j, b_j) \le C_i = (r_i, g_i, b_i)$, if $r_j \le r_i, g_j \le g_i$ and $b_j \le b_i$. In fact, we can describe cumulative color histogram in terms of *crisp* membership matrix $M = [m_{ij}]_{n\times n'}$ and $n' = n$, which is defined as follows:

$$m_{ij} = \begin{cases} 1, & \text{if } C_j \le C_i \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

Therefore,

$$\check{H}_{n\times 1} = M_{n\times n} H_{n\times 1}. \tag{9}$$

For example, given an ordered color histogram with eight bins and $C_i < C_j$, where $i < j$, the membership matrix $M_{8\times 8}$ of cumulative color histogram is

$$M_{8\times 8} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}. \tag{10}$$

Note that inherently cumulative color histogram also considers the color similarity across all color bins. However, FCH is more general as its *fuzzy* (rather than *crisp*) membership matrix can be adjusted according to different noise interference and applications.

Quadratic histogram distance [1] provides more stable and consistent matching measurement than other similarity measures between two CCHs. Given two color images Q and T, the quadratic distance between their n-bin CCHs, H_Q and H_T, is given by

$$d_q^2(H_Q, H_T) = [H_Q - H_T]_{n\times 1}^T A_{n\times n} [H_Q - H_T]_{n\times 1} \tag{11}$$

where $A = [a_{ij}]_{n\times n}$ is a weighted similarity matrix and a_{ij} denotes perceptual similarity between color bins i and j. With a suitable membership matrix $M = [m_{ij}]_{n\times n'}$, the FCHs of images Q and T can be computed by (4), respectively. On the other hand, the squared Euclidean distance between their n-bin FCHs is

$$\begin{aligned} d_E^2(F_Q, F_T) &= [F_Q - F_T]_{n\times 1}^T [F_Q - F_T]_{n\times 1} \\ &= [H_Q - H_T]_{n'\times 1}^T M_{n\times n'}^T M_{n\times n'} [H_Q - H_T]_{n'\times 1} \\ &= [H_Q - H_T]_{n'\times 1}^T A_{n'\times n'} [H_Q - H_T]_{n'\times 1}. \end{aligned} \tag{12}$$

Compared with (11), the simple squared Euclidean distance between two n-bin FCHs is equivalent to the quadratic histogram distance between their n'-bin CCHs. Note that the computationally intensive matrix multiplication in computing quadratic distance of CCHs (11) is incurred at *online* retrieval stage. On the other hand, our FCH-based representation simply applies Euclidean distance measurement, and the matrix multiplication is desirably avoided at *online* retrieval stage, because it has already been performed in the *offline* indexing stage according to (4).

From (12), it also shows that our FCH-based measurement could preserve more detailed color similarity information than CCH-based quadratic distance measurement with the same number of histogram bins because $n \ll n'$. This indicates that it is possible to exploit FCH with fewer number of histogram bins to efficiently represent color distribution than CCH.

VI. Experimental Results of Image Retrieval

A. Retrieval Performance Evaluation Criterion

We evaluate the performance of image retrieval according to *normalized rank sum* (NRS) [22], which is defined as follows. From a manually predefined target image set $\{I_t\}$ containing n_t similar images stored in the database, a query image $I \in \{I_t\}$ is selected for performing image retrieval experiments. If all the images in the database were sorted according to the similarity measured with respect to query image I, the *rank* of each image corresponds to its location in the sorted list. When all the n_t images in the target image set $\{I_t\}$ appear in the first n_t locations in the sorted list, the ideal (or best) retrieval performance is achieved. The *rank sum* of the query image I, which is defined as the sum of the ranks of all the n_t target images (i.e., the denominator of (13)), denotes the performance of a retrieval method exploited. To compare the rank sums of target image sets with different set sizes, the NRS of image I is required and defined as

$$\mathrm{NRS}(I) = \frac{n_t(n_t+1)/2}{\sum_{l=1}^{n_t} \mathrm{rank}(I_l)}. \tag{13}$$

Note that the rank sum in the denominator is normalized by $n_t(n_t+1)/2$ in the numerator—the rank sum when the retrieval

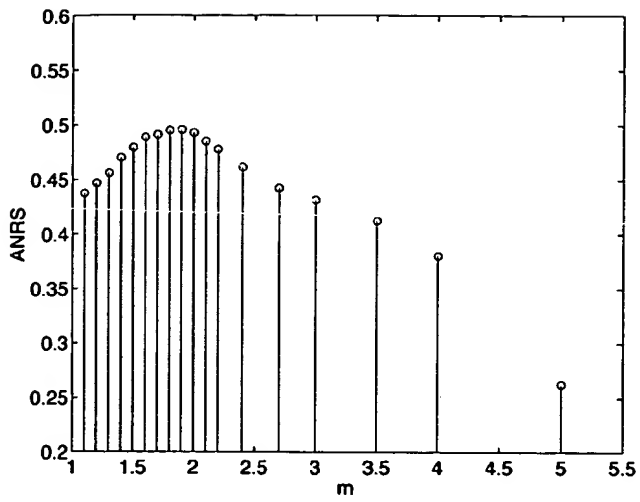

Fig. 2. ANRS values of image retrieval using FCHs with 18 different weighting exponents empirically determined, i.e., $m = 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.4, 2.7, 3.0, 3.5, 4.0$, and 5.0, respectively. (ANRS' range [0, 0.2] is omitted for the purpose of presentation.)

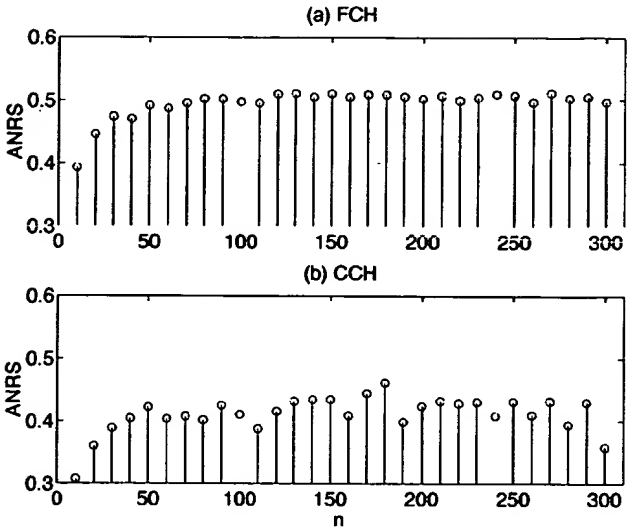

Fig. 3. ANRS values of image retrieval using (a) FCHs ($m = 1.9$) and (b) CCHs with bin numbers $n = 10, 20, 30, \ldots, 290$, and 300, respectively. (ANRS' range [0, 0.3] is omitted for the purpose of presentation.)

performance is *ideal* as described earlier. As the NRS value is approaching one, it indicates that the retrieval performance is getting close to ideal. This makes the NRS measurement independent of the size of the target image set $\{I_t\}$. Also note that between two consecutive correctly retrieved images lie average $\lceil 1/\mathrm{NRS} \rceil - 1$ incorrectly retrieved images [23].

B. FCH Parameter Selection

In order to evaluate the performance of FCH representation exploited in image indexing and retrieval, we establish an image database containing about 500 color images with various sizes and a wide range of image content, such as nature scenes, animals, buildings, etc.

Our experiments for determining FCH parameter were carried out based on global color distribution of the entire image. For that, we selected 39 target sets from our image database based on their global color distribution. Each target image set contains a set of images having similar main object and background, but with some variations in position, viewing angle, illumination, etc.

According to the scheme on computing FCH as described in Section IV, we first uniformly quantize the given RGB color space into $n' = 16^3 = 4096$ color bins [24]. Thus, the weighting exponent m and bin number n are the two main parameters which jointly influence the performance of FCH-based image retrieval. In our experiments, we empirically chose 18 values of m and 30 values of n as shown in Figs. 2 and 3. With each of the 18×30 parameter combinations, the membership matrix was obtained using FCM algorithm, and the FCHs of all the database images were computed. Each image contained in the 39 target image sets was selected as the query image, and the NRS value of the query image was computed by using the Euclidean distance as the similarity measurement. Then, the *average* NRS (ANRS) value over the entire image database was computed. The ANRS value thus represents the performance of image retrieval based on FCHs with the given parameter combination—the larger the ANRS value, the better the parameter combination. For comparison, the ANRS values for CCHs were also obtained in the same way.

With the 18×30 parameter compositions, 18×30 ANRS values of FCH-based retrieval were obtained. We first determine the optimal m value as follows. For each of the 18 values, we averaged all the 30 ANRS values with different n values. The obtained 18 ANRS values are shown in Fig. 2. It suggests that the choice of $m = 1.9$ achieves the best retrieval performance in our experiments.

The 30 ANRS values of FCH-based retrieval with $m = 1.9$ under different n values are shown in Fig. 3(a), and the corresponding 30 ANRS values of CCH-based retrieval are shown in Fig. 3(b). Comparing these two subfigures, we can see that the retrieval performance using FCHs is better than the performance using CCHs under the same bin number. Moreover, it also indicates that the FCH-based image retrieval is less sensitive to the bin number changes. As the quantization errors are intimately related to the bin number used, these results demonstrate that FCH is more robust (i.e., less sensitive) to quantization errors than CCH.

C. Retrieval Sensitivity Under Lighting Intensity Changes

To study the robustness of FCH with respect to lighting intensity changes, we carry out the following image retrieval experiments. First, we select an image from the database as the query image. Then, the query image is processed by using, say, Photoshop to create ten images under lighting intensity changes with amount varying from $-25, -20, -15, -10, -5, +5, +10, +15, +20$ to $+25$, respectively. These ten images are then added back to the database. For comparison, both FCH (using $m = 1.9$) and CCH for database images are independently computed with 64 bins (i.e., $n = 64$) each. Finally, all the database images are sorted with

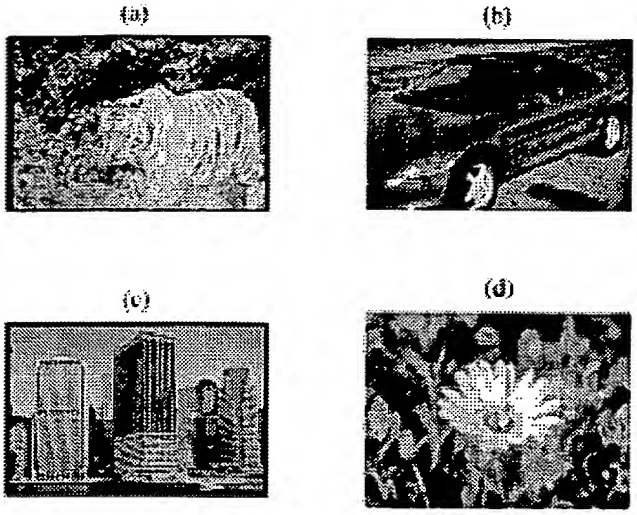

Fig. 4. Arbitrarily selected four query images for sensitivity studies under lighting intensity changes.

TABLE I
RANKS OF CORRESPONDING TEN PROCESSED IMAGES UNDER VARIOUS DEGREES OF LIGHTING INTENSITY CHANGES WITH RESPECT TO EACH QUERY IMAGE AS SHOWN IN FIG. 4

query image	feature	ranks of the corresponding 10 images									
(a)	FCH	4	6	10	16	17	23	29	36	69	141
	CCH	4	5	10	17	18	29	42	46	128	239
(b)	FCH	2	3	4	5	6	7	10	22	45	143
	CCH	2	3	5	6	9	12	18	22	121	390
(c)	FCH	3	4	5	6	7	8	10	14	19	31
	CCH	3	5	8	9	11	16	38	116	124	189
(d)	FCH	2	3	4	5	6	7	8	14	74	137
	CCH	2	3	4	5	6	7	16	18	91	370

respect to the query image based on the Euclidean distance measurement.

Four query images are arbitrarily selected from our image database as presented in Fig. 4. The experimental results are documented in Table I, which shows the ranks of the corresponding ten processed (i.e., under lighting intensity changes as previously mentioned) images. For the purpose of presentation, note that the entries of each row have been arranged from high to low in their ranks without considering their corresponding lighting intensity changes individually. The justification is quite clear that as long as the images from the target image set could be retrieved, the exact ordering among themselves is not important anymore. It clearly demonstrates that the ranks obtained by FCH are much higher than those obtained by CCH. Similar results and conclusion are also obtained from extended simulation experiments using other database images as the query images, respectively. Therefore, our proposed FCH is more robust to lighting intensity changes than CCH for the task of image indexing and retrieval.

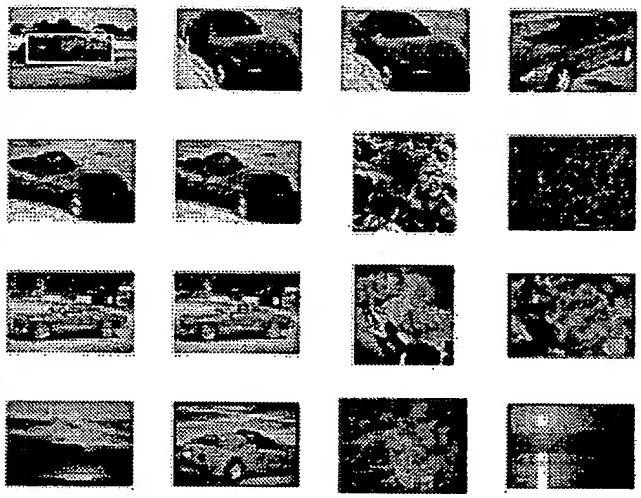

Fig. 5. Top-left image is the query image with a user-selected local region indicated by the white-line bounding box. With respect to the selected local region, the 16 most similar images retrieved from a database containing about 500 color images. The retrieval criterion is based on the Euclidean distance between FCHs.

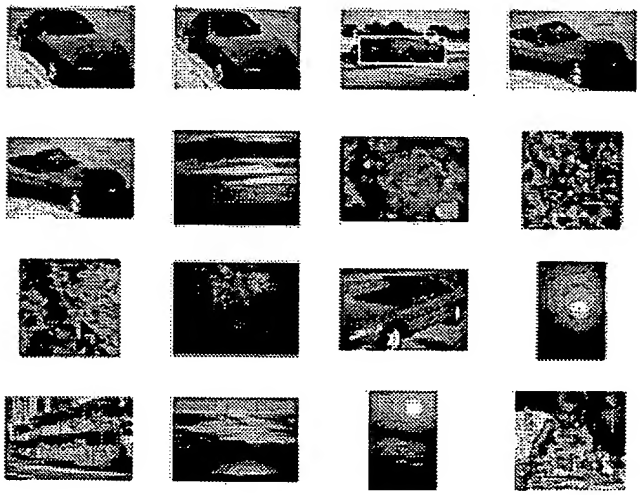

Fig. 6. Same experiments as those of Fig. 5 are conducted by exploiting CCHs. The retrieval criterion is based on the Euclidean distance between CCHs.

D. Regional Image Retrieval

Image indexing by localized or regional color distribution provides partial or subimage matching between images. For example, if the user is interested in finding all the images containing human faces regardless their backgrounds, the regional indexing approach would be more effective as the background information will be completely excluded for similarity matching. For that, we employ the hierarchical partition scheme proposed in Dimai's work [23] in our experiment.

For region-based image retrieval, the query object selected by the user from the query image should be matched by those database images that contain such object but appearing at different locations with possibly variable sizes and angles. To achieve this goal, we systematically partition each database image into subimages in order to increase the chances of

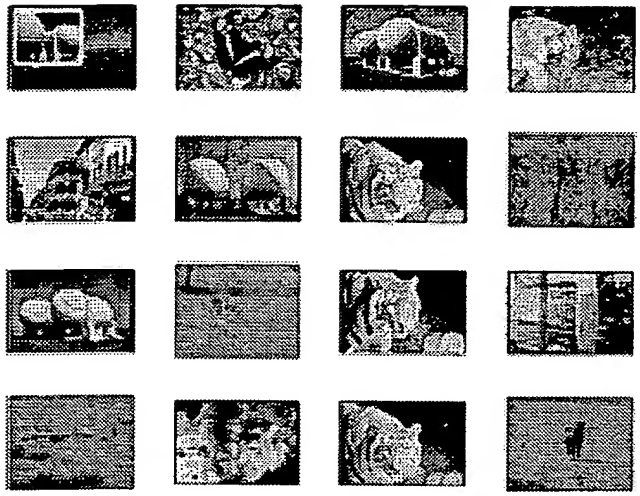

Fig. 7. Another retrieval result based on FCHs.

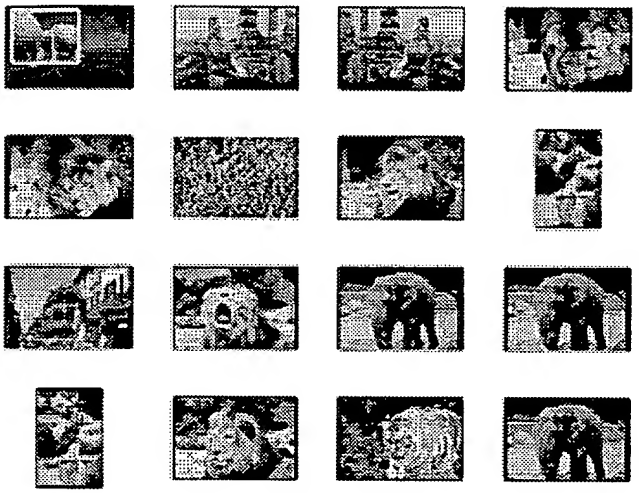

Fig. 8. Same experiments as those of Fig. 7 are conducted by exploiting CCHs. Note that the three target images as presented in Fig. 7 fail to be retrieved.

matching the query object from the query image. The methodology of dividing each database image into three hierarchical levels as introduced in [23] is adopted here and generates overlapping subimages. The highest level is the image itself, and the image is then equally portioned into 3×3 overlapping rectangle regions in the second level which has each side length be a half of the corresponding side length of the original image. Similarly, with finer partition, the lowest level is composed by 5×5 rectangle regions with each region having its side lengths being one-third of the image side lengths, respectively. Therefore, total 35 $(=1 + 9 + 25)$ rectangular regions are obtained for each database image.

For each of 35 regions, its 64-bin FCH (with $m = 1.9$) and 64-bin CCH are computed as the feature vector, respectively. We also employ the Euclidean distance as the similarity measure for both cases. The similarity between the query local image and a database image is measured by the minimum distance between the feature vectors of the query local image and all the 35 rectangle subregions of each database image.

TABLE II
RETRIEVAL PERFORMANCE (NRS VALUE) FOR 50 CCQS ON CCD USING FCH AND CCH, RESPECTIVELY

CCQ	CCH	FCH	CCQ	CCH	FCH
flower garden	0.0185	0.3846	quiz scene	0.3030	1
rock and sky	0.0053	0.2586	speaker	1	1
news anchor	0.9483	1	man and horse	0.9545	1
walking people	0.2122	0.6892	space earth	1	1
baldheaded man 1	0.9740	1	fountain	0.0473	0.1986
sports reporters	0.0366	0.2500	graphics before news	0.5056	1
congress	0.1023	0.3600	Ron Reagan	0.8182	0.5921
baldheaded man 2	0.1400	0.9215	basketball game overlay	1	1
castle	1	1	glass roof	0.0383	0.0319
black clothes lady	0.7241	1	snow clad mountain	0.1000	0.3030
singer	1	0.8824	outdoor/boats	0.0139	0.0345
strange hair	0.1685	0.8333	by the water	0.2000	0.6250
leather jacket people	0.4472	0.3846	couple	0.0142	1
man with placard	0.0439	0.4839	shop	0.0095	0.1556
people on the red	0.5769	0.2830	flower(indoor)	0.0156	0.1639
snake	1	0.9873	playing on the street	0.0493	0.1899
fish	1	1	road with trees/grass	0.0047	0.0481
tapirs	1	1	children/rock/grass	0.2128	0.0144
butterfly	1	1	Asian building	0.0044	0.0167
small monkey	0.9512	0.7723	containers	0.0041	0.2381
landscape image 1	0.1370	0.2381	sunset over lake	0.0203	0.0285
landscape image 2	0.2778	0.0323	big pipes	0.0321	0.0722
landscape image 3	1	1	man in white shirt	0.0084	0.0458
indoor image	0.1608	0.3035	wooden shack	0.0463	0.2727
anchor person	0.8407	0.9563	ruins	0.0059	0.0098

Our experimental results show that the performance of regional image retrieval by FCH is consistently better than or equivalent to that by CCH in general. Two examples are presented in Figs. 5 and 6 for demonstration. The first (i.e., top-left) picture in Fig. 5 is the query image with a selected local region indicated by the white-line bounding box imposed by the user. The images presented in Figs. 5 and 6 in the order of ranking show the retrieval results after searching for those images containing a "red car" from the database based on FCH's and CCH's representation, respectively. In Fig. 5, the 16 most similar images retrieved based on FCHs include all the 9 images containing a red car. Note that the last three "red car" images in Fig. 5 do not appear in the 16 most similar images retrieved by exploiting CCHs as shown in Fig. 6. Note that even the query image itself is not being ranked as the most relevant retrieval in Fig. 6, as it normally should be.

Another set of retrieval results are shown in Figs. 7 and 8. Note that the three target images in Fig. 7 (with ranking of third, sixth, and ninth, respectively) are failed to be retrieved in Fig. 8.

E. Image Retrieval Results on MPEG-7 Testing Database

Common color dataset (CCD) is established in MPEG-7 as the test database for conducting color core experiments [25]. Among the 5466 images contained in this database,

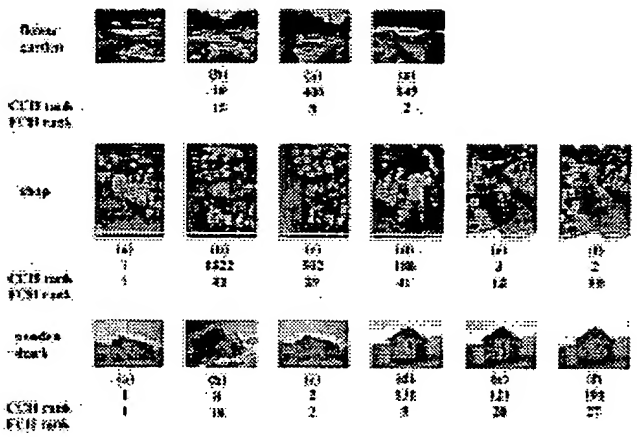

Fig. 9. Retrieval results of using three CCQs as shown in images (a) in each rows, based on FCH and CCH, respectively. The images (a)–(f) in each row are the corresponding GTS with respect to the query image (a) in the same row.

50 *common color queries* (CCQs) and their corresponding so-called *ground truth sets* (GTSs) are defined for the purpose of image retrieval based on color. As mentioned in Section IV, the value of m should be adjusted according to different image retrieval applications. Our experimental results indicate that FCH with $m = 1.2$ achieves best retrieval performance for these queries on CCD. Here, both FCH and CCH are computed based on global color distribution of the entire image. Table II documents the retrieval performance for these queries using 64-bin FCH ($m = 1.2$) and 64-bin CCH, respectively. Note that FCH achieves better performance than CCH in most cases. Among these CCQs, three retrieval results are shown in Fig. 9. Comparing these results, we can see that FCH is less sensitive to noisy interference (i.e., small scene changes and illumination changes) than CCH experimented on this database.

VII. Conclusion

In this paper, we introduce a novel descriptor on representing color images, called *fuzzy color histogram* (FCH), with mathematical development. For computing FCHs, we propose an efficient method based on fuzzy c-means clustering algorithm performed on the color components recorded in the perceptually uniform CIELAB color space. Note that our proposed FCH is generic and can be directly employed in other color spaces and for various application fields as well. Experimental results show that our FCH is less sensitive and more robust than CCH on dealing with lighting intensity changes, quantization errors, region-of-interest image retrieval, and possibly other uncovered aspects in new applications.

From the observation of the interplay between FCH and quadratic histogram distance, our proposed FCH not only addresses the noise sensitivity issue of CCH but also avoids intensive online computation encountered in computing the quadratic histogram distance. The preliminary results of this work proposed to MPEG-7 [26] since the *retrieval sensitivity* was recognized as an indispensable issue that needs to be satisfactorily addressed.

Finally, exploiting FCH into other image processing frameworks and even extending similar soft clustering approach to other low-level visual features (e.g., shape, texture, etc.) are also recommended here.

Acknowledgment

The authors would like to thank the anonymous reviewers for their helpful comments which improved the quality of this paper.

References

[1] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Querying by image and video content: The QBIC system," *IEEE Trans. Comput.*, vol. 28, pp. 23–32, 1995.

[2] J. R. Smith and S.-F. Chang, "VisualSEEK: A fully automated content-based image query system," in *Proc. ACM Multimedia*, Nov. 1996, pp. 87–98.

[3] W. Y. Ma and B. S. Manjunath, "NETRA: A toolbox for navigating large image databases," in *Proc. IEEE ICIP*, 1997, pp. 925–928.

[4] M. J. Swain and D. H. Ballard, "Color indexing," *Int. J. Comput. Vis.*, vol. 7, no. 1, pp. 11–32, 1991.

[5] J. R. Smith and S.-F. Chang, "Single color extraction and image query," in *Proc. ICIP*, vol. 3, 1995, pp. 528–531.

[6] B. V. Funt and G. D. Finlayson, "Color constant color indexing," *Pattern Anal. Machine Intell.*, vol. 17, no. 5, pp. 522–529, 1995.

[7] M. S. Drew, J. Wei, and Z.-N. Li, "Illumination-invariant color object recognition via compressed chromaticity histograms of color-channel-normalized images," in *Proc. 6th Int. Conf. Computer Vision*, Jan. 1998, pp. 533–540.

[8] M. Stricker and M. Orengo, "Similarity of color images," *Proc. SPIE*, vol. 2420, pp. 381–392, 1995.

[9] J. Hafner, H. S. Sawhney, W. Equitz, M. Flicker, and W. Niblack, "Efficient color histogram indexing for quadratic form distance functions," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, pp. 729–736, July 1995.

[10] M. K. Mandal, T. Aboulnasr, and S. Panchanathan, "Image indexing using moments and wavelets," *IEEE Trans. Consumer Electron.*, vol. 42, pp. 557–565, Aug. 1996.

[11] J. R. Smith and S. F. Chang, "Tools and techniques for color image retrieval," *Proc. SPIE*, vol. 2670, pp. 1630–1639, Feb. 1996.

[12] M. Stricker and A. Dimai, "Spectral covariance and fuzzy regions for image indexing," *Mach. Vis. Applicat.*, vol. 10, pp. 66–73, 1997.

[13] U. Gargi and R. Kasturi, "Image database querying using a multi-scale localized color representation," in *Proc. IEEE Workshop Content-Based Access of Image and Video Libraries*, June 1999, pp. 28–32.

[14] H. Yamamoto, H. Iwasa, N. Yokoya, and H. Takemura, "Content-based similarity retrieval of images based on spatial color distributions," in *Proc. Int. Conf. Image Analysis and Processing*, 1999, pp. 951–956.

[15] G. Pass and R. Zabih, "Histogram refinement for content-based image retrieval," in *Proc. 3rd IEEE Workshop Applications Computer Vision*, Dec. 1996, pp. 96–102.

[16] B. Y. Kim, H. J. Kim, and S. J. Jiang, "Image retrieval based on color coherence," in *Proc. TENCON 99*, vol. 1, 1999, pp. 178–181.

[17] J. Huang, S. R. Kumar, M. Mitra, W. J. Zhu, and R. Zabih, "Image indexing using color correlograms," in *Proc. IEEE CVPR*, 1997, pp. 762–768.

[18] J. Huang, S. R. Kumar, M. Mitra, and W.-J. Zhu, "Spatial color indexing and applications," in *Proc. 6th Int. Conf. Computer Vision*, Jan. 1998, pp. 602–607.

[19] B. Hill, T. Roger, and F. W. Vorhagen, "Comparative analysis of the quantization of color spaces on the basis of the CIELAB color-difference formula," *ACM Trans. Graph.*, vol. 16, no. 2, pp. 109–154, Apr. 1997.

[20] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum, 1981.

[21] M. R. Rezaee, B. P. F. LeLieveldt, and J. H. C. Reiber, "A new cluster validity index for the fuzzy c-means," *Pattern Recognit. Lett.*, vol. 19, pp. 237–246, 1998.

[22] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz, "Efficient and effective querying by image content," *J. Intell. Inform. Syst.*, vol. 3, pp. 231–262, July 1994.

[23] A. Dimai, "Differences of global features for region indexing," Swiss Federal Inst. Technol., Lausanne, Tech. Rep. 177, 1997.
[24] J. Wang, W. J. Yang, and R. Acharya, "Color clustering techniques for color-content-based image retrieval from image databases," in *Proc. IEEE Int. Conf. Multimedia Computing and Systems*, 1997, pp. 442–449.
[25] D. Zier and J.-R. Ohm, "Common dataset and queries in MPEG-7 color core experiments [Doc. M5060]," presented at the 49th MPEG Meeting (ISO/IEC/JTC1/SC29/WG11), Melbourne, Australia, Oct. 1999.
[26] J. Han and K.-K. Ma, "A novel color histogram representation for color images [Doc. M5510]," presented at the 50th MPEG Meeting (ISO/IEC/JTC1/SC29/WG11), Maui, HI, USA, Dec. 1999.

Ju Han received the B.S. degree from Shandong University, China, in 1994, the M.S. degree from the Institute of Automation, Chinese Academy of Sciences, in 1998, both in electrical engineering, and the M.Eng. degree in electrical and electronic engineering from Nanyang Technological University, Singapore, in 2000. He is currently pursuing the Ph.D. degree in the Department of Electrical Engineering, University of California, Riverside.

His research interests include image/video indexing and retrieval, automatic gait recognition, and other image processing and computer vision related areas.

Kai-Kuang Ma (S'80-M'84-SM'95) received the Ph.D. degree from North Carolina State University, Raleigh and the M.S. degree from Duke University, Durham, NC, both in electrical engineering, and the B.E. degree from Chung Yuan Christian University, Chung-Li, Taiwan, R.O.C., in electronic engineering.

He is presently an Associate Professor with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. Prior to this, he was with the Institute of Microelectronics (IME), National University of Singapore (1992–1995), IBM Corporation, Kingston, NY, and then Research Triangle Park, NC (1984–1992). His research interests are in the areas of multimedia signal processing and communications, including digital image/video coding, content-based image/video indexing and retrieval, video-object segmentation, wavelets and filter banks, joint source and channel coding for robust visual communications, nonlinear denoising filter, error concealment and artifact postprocessing, clustering and pattern recognition, and multimedia networking and quality of service. He is an Associate Editor of the *International Journal of Image and Graphics*. He has been serving as committee member in multiple international conferences and paper reviewer of many international journals. From 1997 to 2001, he served as the Chairman and Head of Delegation for Singapore in MPEG and JPEG. On the MPEG contributions, the proposed fast motion estimation algorithms from his research team have been adopted by the MPEG-4 standards. He was serving as the General and Organizing Chair of ISO/IEC JTC1/SC29 Plenary Meetings and a series of Working Group meetings in March 2001.

Dr. Ma is serving as Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS and Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA. He is a committee member of the IEEE Communications Society on Multimedia Communications Technical Committee. He is a Technical Program Co-Chair, IEEE International Conference on Image Processing (ICIP) 2004. He is also serving as the Chairman of IEEE Signal Processing Singapore Chapter. He is a member of Sigma Xi and Eta Kappa Nu.